

The Complexity of Deterministic Source
Encoding with a Fidelity Criterion*

by

John E. Savage

May, 1971

*This work was supported by NASA Grants NGR 40-002-082, NGR 40-002-090
and NSF Grant GK-13162.

The Complexity of Deterministic Source Encoding
with a Fidelity Critereon

by

John E. Savage

Center for Computer and Information Sciences
and
Division of Engineering

Brown University
Providence, Rhode Island 02912
U.S.A.

Abstract

This paper deals with the block encoding of discrete, memoryless sources with equiprobable outputs subject to a fidelity critereon. The encoder is assumed to be deterministic and a lower bound is derived on the amount of computational work required to encode such sources as a function of the code rate and fidelity of the encoding. An example is given to show that the bound cannot be substantially improved. These bounds apply primarily to sources and distortion measures which are such that the minimum distortion given by the rate-distortion bound is not achievable with a finite block length, deterministic encoder.

1. Introduction

Codes are used both for the encoding of sources with a fidelity criterion¹ and for error-correction. In the case of error-correction, decoders are usually far more complex than encoders while in the source encoding case the reverse is true. This difference exists because source encoders which represent source outputs with some distortion and at reduced code rate require a many-to-one map from source sequences to code words, and this map is potentially much more complex than the one-to-one map realized by encoders for error-correcting codes. The objective of this paper is to shed light on the source encoding problem by introducing to it the complexity measure called "computational work" which has been successfully applied to the decoding problem².

In this section we define "computational work," the source encoding problem and introduce models for encoding machines. In Section 2 a lower bound to the computational work required to source encode a discrete, memoryless source (DMS) is derived and a high rate encoder for a binary source is exhibited which encodes with a work near the minimum. In the concluding section the problems associated with the use of decoders for error-correcting codes as source encoders are discussed.

Let the source produce N digits from a source alphabet Z , namely, the vector $\vec{z} = (z_1, \dots, z_N)$. The source encoder maps \vec{z} into $\vec{y} = (y_1, \dots, y_N)$ where $y_i \in Y$, a second alphabet. In general, the map from \vec{z} to \vec{y} is many-to-one and as a result some distortion in the representation of \vec{z} is introduced. We measure this distortion with a function $d(,)$ which

we assume has the following properties*:

$$\begin{aligned} 0 \leq d(z,y) \leq \infty & \quad z \in Z, y \in Y \\ d(z,y(z)) = 0 & \quad \text{some } y(z) \in Y \end{aligned} \quad (1)$$

Then, the distortion between \vec{z} and \vec{y} , $d(\vec{z}, \vec{y})$ is defined by

$$d(\vec{z}, \vec{y}) = \frac{1}{N} \sum_{i=1}^N d(z_i, y_i) \quad (2)$$

The map from \vec{z} to \vec{y} could be stochastic since \vec{y} could be the output of a noisy channel when \vec{z} is the input.

Shannon¹ has shown for the DMS that the minimum average distortion D and the source rate R must satisfy the inequality

$$R \leq R(D) \quad (3)$$

where $R(D)$ is a rate-distortion function defined in terms of the distortion measure $d(,)$. Since Shannon's encoders could contain a stochastic element, it is surprising that (as Goblick⁴ has shown) the same results hold when the map from \vec{z} to \vec{y} is deterministic. Pilc⁵ has derived bounds on the rate of approach to the $R(D)$ bound as a function of encoding length N and finds that stochastic encoders offer a small advantage over deterministic encoders.

The analysis that we shall do is limited by technical considerations to deterministic source encoders. Therefore, we limit our attention to such encoders which can now be defined by a many-to-one function $f: Z^N \rightarrow Y^N$ where if $\vec{z} \in Z^N$ and $\vec{y} \in Y^N$ then

$$\vec{y} = f(\vec{z}) \quad (4)$$

Following Goblick⁴ and Pilc⁵, if there are M encoder outputs \vec{y} , the encoder rate R is defined by

$$R = \frac{\log_2 M}{N} \quad (5)$$

*Pinkston³ has shown that there is no loss of generality in these assumptions.

The source outputs are assumed to be statistically independent and identically distributed according to the distribution $\{Q(i), 1 \leq i \leq |Z|\}$.

Suppose that f has ℓ inputs which are equal to outputs and let these inputs be z_1, z_2, \dots, z_ℓ . Form πf by $\pi f(z_1, \dots, z_N) = (y_{\pi(1)}, \dots, y_{\pi(N)})$ where π is a permutation of $\{1, 2, \dots, N\}$ and $f(z_1, \dots, z_N) = (y_1, \dots, y_N)$. Let P be the set of permutations such that if $\pi \in P$, $y_{\pi(1)} = z_1, \dots, y_{\pi(\ell)} = z_\ell$. Then, we define the average distortion D associated with f by

$$D = \min_{\pi \in P} E\{d(\bar{z}, \pi f(\bar{z}))\} \quad (6)$$

where the expectation is taken over the source ensemble. While this definition is slightly different from that commonly used, the rate-distortion function still applies.

The complexity of an encoding function $f: Z^N \rightarrow Y^N$ will now be defined. Let $h: Z \rightarrow \{0, 1\}^m$ be a 1-1 into map of the set Z and let $\ell: Y \rightarrow \{0, 1\}^n$ be a 1-1 into map of Y . Then, $f^*: [\{0, 1\}^m]^N \rightarrow [\{0, 1\}^n]^N$, obtained from f by composition with functions h and ℓ , is a binary representation of f .

Given a binary function $g: \{0, 1\}^p \rightarrow \{0, 1\}^q$ we say that g is computed by a combinational machine (a directed, acyclic graph) with primitives Ω (set of Boolean functions) if such a machine accepts as inputs the variables of g and produces the value of g at specified points in the machine. The primitives Ω are the basic set of operations permissible and might consist of the 2-input AND, the 2-input OR and the NOT functions, for example. If no element of Ω has more than r inputs, Ω is said to have fan-in of r . The combinational complexity of g , $C_\Omega(g)$, is the smallest number of primitives in any combinational machine realizing g . We extend the definition to non-binary encoding functions f by defining $C_\Omega(f)$ as the

*Each occurrence of z and y may have a different encoding function h or ℓ .

minimum of $C_{\Omega}(f^*)$ over all encodings of \bar{z} and \bar{y} .[†]

A sequential machine $S = \langle S, I, \delta, \lambda, 0; T \rangle$ has a finite state set S , input alphabet I and output alphabet O . It executes T cycles and has state transition function $\delta: SXI \rightarrow S$ and output function $\lambda: S \rightarrow O$. Then, $C_{\Omega}^*(S)$ is defined as the combinational complexity of δ, λ subject to the restriction that each occurrence of the set S be given the same encoding. S is said to compute $\lambda^{(n)}: S \times I^{n-1} \rightarrow O$ as defined by: $\lambda^{(1)} = \lambda$, $\delta^{(1)} = \delta$, $\delta^{(n)}: S \times I^n \rightarrow S$

$$\begin{aligned} \delta^{(n)}(s; y_1, y_2, \dots, y_n) &= \delta(\delta^{(n-1)}(s; y_1, \dots, y_{n-1}), y_n) \\ \lambda^{(n)}(s; y_1, \dots, y_{n-1}) &= \lambda(\delta^{(n-1)}(s; y_1, \dots, y_{n-1})) \end{aligned} \quad (7)$$

Then, if $C_{\Omega}(\lambda^{(1)}, \dots, \lambda^{(t)})$ is the combinational complexity of the functions computed by S then it is easily shown^{2,6} that

$$C_{\Omega}(\lambda^{(1)}, \lambda^{(2)}, \dots, \lambda^{(T)}) \leq C_{\Omega}^*(S) T \quad (8)$$

Theorem 1 Let $f: Z^N \rightarrow Y^N$ be computed by $S = \langle S, I, \delta, \lambda, 0; T \rangle$.

Then,

$$C_{\Omega}(f) \leq C_{\Omega}^*(S) T \quad (9)$$

If f is computed by S then $C_{\Omega}(f) \leq C_{\Omega}(\lambda^{(1)}, \dots, \lambda^{(T)})$ and the inequality follows. This theorem can be extended to a collection of interconnected sequential machines which compute f .

We call $W = C_{\Omega}^*(S) T$ the computational work performed by S because it is the equivalent number of logical operations which S executes. Furthermore, it can be shown⁶ that $C_{\Omega}^*(S)$ is proportional to the storage capacity

[†]Each occurrence of Z and Y may have a different encoding function.

of S when S consists of a machine with a large amount of random access or tape storage. The theorem then states that a source encoding with f is possible only if an amount of computational work at least as large as $C_{\Omega}(f)$ is completed. We show in the next section that $C_{\Omega}(f)$ must be large for many sources and distortion measures if the source is to be encoded at rate R and distortion D near that given by the rate distortion function.

2. Bounds on Computational Work.

We now develop a lower bound to $C_{\Omega}(f)$. Let π^* be the minimizing permutation in (6). Then, if f is dependent on N' of its inputs, so is π^*f and

$$C_{\Omega}(f) = C_{\Omega}(\pi^*f) \geq \frac{N' - \ell}{r} \quad (10)$$

where r is the fan-in of Ω . This follows because (1) no output is equal to two inputs (outputs can only be connected together through logic elements) (2) if some input variable of f is equal to an output variable, this input is one of the ℓ inputs and (3) the $N' - \ell$ remaining inputs are encoded into binary variables and f^* must depend on at least one of the binary variables for each of the inputs to f . These binary inputs must be applied to a logic element and each logic element has at most r inputs. The next step is to lower bound $N' - \ell$.

We observe that the number of points in Z^N on which f depends, namely, $|Z|^{N'}$, must be at least as large as M , the size of the range of f . Therefore,

$$N' \geq NR / \log_2 |Z|. \quad (11)$$

Write $\bar{z} = (\bar{v}, \bar{w})$ where \bar{v} represents the first ℓ components of \bar{z} and \bar{w} the remaining $N - \ell$ components. Then, $\pi^*f(\bar{z})$ can be written as $(\bar{v}, f_{\bar{v}}(\bar{w}))$. Let $M(\bar{v})$ be the size of the range of $f_{\bar{v}}(\bar{w})$. Then,

$$\sum_{\bar{v}} |Z|^{-\ell} M(\bar{v}) = M/|Z|^{\ell} \quad (12)$$

Also, if the source is a statistically independent letter source, the probability $p(\bar{z}) = p_v(\bar{v})p_w(\bar{w})$ and since $d(z_i, y_i) \geq 0$ we have

$$D \geq \frac{N-\ell}{N} \sum_{\bar{v}} p_v(\bar{v}) E_w \{d(\bar{w}, f_{\bar{v}}(\bar{w}))\} \quad (13)$$

We call $D(\bar{v}) = E_w \{d(\bar{w}, f_{\bar{v}}(\bar{w}))\}$ and if the source outputs are equiprobable, then $p_v(\bar{v}) = |Z|^{-\ell}$ and

$$\sum_{\bar{v}} |Z|^{-\ell} D(\bar{v}) \leq D \frac{N}{N-\ell} \quad (14)$$

In Appendix A it is shown that there exists a \bar{v} such that

$$M(\bar{v}) \leq 2M/|Z|^{\ell} \quad (15a)$$

$$D(\bar{v}) \leq 2 D \frac{N}{N-\ell} \quad (15b)$$

where $M(\bar{v})$ and $D(\bar{v})$ are the number of code words and average distortion associated with $f_{\bar{v}}: Z^{N-\ell} \rightarrow Y^{N-\ell}$.

Taking the base 2 logarithm of both sides of (15a) and letting $R(\bar{v}) = (\log_2 M(\bar{v})) / (N-\ell)$ we have

$$\ell \leq N[1/N + R - R(\bar{v})] / [\log_2 |Z| - R(\bar{v})] \quad (16)$$

From (15b) and the fact that $D(\bar{v}) \geq R^{-1}(R(\bar{v}))$ we have

$$L \geq N[1 - 2D/R^{-1}(R(\bar{v}))] \quad (17)$$

and combining these two equations we have

$$[1 - 2D/R^{-1}(R(\bar{v}))][\log_2 |Z| - R(\bar{v})] + R(\bar{v}) < 1/N + R \quad (18)$$

In Appendix B it is shown that if $R(D)$ has a continuous first derivative then the left-hand function is decreasing in increasing $R(\bar{v})$ and $R(\bar{v}) > 0$ if the following inequality is satisfied:

$$R < (1 - 2D/D_{\max})R_{\max} - 1/N \quad (19)$$

where D_{\max} is the maximum average distortion and $R_{\max} = \log_2 |Z|$ is the largest possible source rate. The locus $(1 - 2D/D_{\max})R_{\max}$ vs D is shown in Figure 1 along with $R(D)$ for a typical source and distortion measure.

Substituting (11) and (16) into (10) we have the following

Theorem 2 Consider a discrete, memoryless and equiprobable letter source and let a distortion measure be given which generates the rate-distortion function $R(D)$ assume that $R(D)$ has a continuous first derivative. Let $R_{\max} = \log_2 |Z|$ be the maximum encoding rate and let D_{\max} be the maximum average distortion. If the source is encoded with average distortion D at rate R by a deterministic encoder realizing a function $f: Z^N \rightarrow Y^N$, then the computational work W which is required satisfies

$$W \geq \frac{N}{r} \alpha(R, D, N) \quad (20)$$

where $\alpha(R, D, N) > 0$ if $N > N_0(R, D)$ and

$$R(D) < R < (1 - 2D/D_{\max})R_{\max} \quad (21)$$

Also, $\lim_{N \rightarrow \infty} \alpha(R, D, N) = \beta(R, D)$ and $\beta(R, D) > 0$ if (21) is satisfied.

Proof The substitution of (11) and (16) into (10) gives

$$C_{\Omega}(f) > \frac{N}{r} \left[\frac{R(\vec{v}) (R_{\max} - R) - 1/N}{R_{\max} (R_{\max} - R(\vec{v}))} \right] = \frac{N}{r} \alpha(R, D, N) \quad (22)$$

But (18) implies that $R(\vec{v}) > 0$ if (14) is satisfied which it is by the assumption (21) for large N . Therefore, for sufficiently large N , $\alpha(R, D, N) > 0$. Q.E.D.

We now show by example that we cannot improve substantially upon the bound of (20). Consider a source encoder for a binary, equiprobable, memoryless letter source which is the minimum distance decoder of a Hamming code. Every binary sequence \vec{z} is at Hamming distance at most one from a code word \vec{y} . Let the distortion measure by the Hamming metric. Then $R(D) = 1 - H(D)$, where $H(D)$ is the entropy of a binary source with letter probabilities of D and $1-D$, and $R_{\max} = 1$, $D_{\max} = 1/2$ and (21) becomes $R < 1 - 4D$. A Hamming code of block length N has rate $R(N) = 1 - \log_2(N+1)/N$ and the average distortion using this code and decoder $D(N)$ satisfies

$$D(N) = \frac{1}{N} 2^{NR(N)} ((N+1)2^{-N}) = \frac{1}{N} \quad (23)$$

Hence, for $N > 15$, the inequality $R < 1 - 4D$ of (21) is satisfied.

The decoder of a Hamming code calculates a syndrome and then identifies the bit presumed to be in error by equating the syndrome with some row of the code parity-check matrix. It has been shown⁷ that the decoder does a work W bounded above by a quantity which approaches $3N^2 \log_2 N$ when 2-input primitives are used. Thus, the lower bound of (10) cannot be substantially improved for large N at large encoding rates.

We now turn to the further bounding of computational work, for which we make use of bounds derived by Pilc⁵. We assume that the source is discrete and memoryless characterized by the letter probabilities $\{p_1, \dots, p_I\}$ over the source alphabet Z . A distortion measure is given and a rate-distortion function $R(D)$ is defined in terms of the semi-invariant moment generating function $\mu(S, p, g)$

$$\mu(S, p, g) = \sum_{i=1}^I p_i \log_2 \left\{ \sum_{j=1}^J g_j e^{S d_{ij}} \right\} \quad (24)$$

where $\{g_i, -g_j\}$ are probabilities over the encoder output alphabet Y and $\{d_{ij}\}$ is the set of values assumed by the function. Then, $S = S_0$ is chosen such that

$$D = \min_g \mu'(S_0, p, g) \quad (25)$$

and we have

$$R(D) = [S_0 \mu'(S_0, p, g_0) - \mu(S_0, p, g_0)] \quad (26)$$

where g_0 is the minimizing probability vector g . Pilc shows that subject to an approximation

$$D \geq D_R + \frac{1}{2|S_0|} \frac{\log_2 N}{N} [1 + o(1)] \quad (27)$$

where $o(1)$ decreases to zero with increasing N . The approximation is explained on page 841 of [5] and has been shown to be unnecessary when the source is doubly uniform, that is when the source letters are equi-probable and the matrix $\{d_{ij}, 1 < i < I, 1 < j < J\}$ is such that rows and columns are permutations of a given row and column and $I = J$. These conditions hold for the source and distortion function in the example given above.

Now solving (29) for large N , that is for D near D_R , where $R(D_R) = R$ we have

$$N \geq \frac{-\log_2 2|S_o|(D - D_R)}{2|S_o|(D - D_R)} \quad (28)$$

where both S_o and D_R are implicit functions of the encoding rate R . Since Pilc's bounds apply strictly to doubly uniform sources, we restrict our principal result to this case. It should be noted that Pilc's lower bounds apply to the distortion measure defined in (6).

Theorem 3 Consider a doubly uniform source with rate-distortion function $R(D)$. Let D_{\max} , and R_{\max} be the maximum average distortion and maximum encoding rate and define D_R by $R(D_R) = R$. Let the source be encoded with a deterministic encoder with average distortion D and rate R . Then, the computational work required to encode the source satisfies

$$W \geq \frac{\left[-\log_2 |S_o|(D - D_R) \right] \alpha'(R, D)}{2|S_o|(D - D_R)} \quad (29)$$

for $D_R < D < D_R + \epsilon$ when $0 < \epsilon \ll 1$ and $\alpha'(R, D) > 0$ if R satisfies (21). Here $S_o = S_o(R)$ is the parameter defined by (25) and it is assumed to be bounded.

Proof $\alpha'(R, D)$ is the function $\alpha(R, D, N)$ with N replaced by the bound of (28). Clearly, for D near D_R , N is large and Theorem 2 applies. Q.E.D.

3. Conclusions

The central objective of source encoding with a fidelity critereon is to find an encoder which provides a distortion near the minimum with a small computational work. One might expect that decoders for error correcting codes could be used since they partition their input spaces into disjoint sets,

as do source encoders. This is an unrealized expectation primarily because those known decoding procedures which decode long codes of non-zero rates with a modest computational work also decode "with uncertainty". That is, they make no decision on many input words; this is true of "bounded distance" decoders. Unfortunately, these uncertainty sets occupy nearly all of the volume of the space of input words so that nearly all inputs result in uncertainty. This is not a problem for error-correction since the received sequences are concentrated with high probability outside this set.

In general, it is a problem for source encoders because in this case all sequences are often equiprobable which means that either no action can be taken on most source sequences or if the same decision is always made, then, the average distortion will be very large.

The search for good source encoders with a fidelity criterion goes on. There is some comfort in the fact that success here is likely to materially aid in the quest for error-correcting decoders which decode with small computational work.

Appendix A

Lemma Let $A(i), B(i) \geq 0$, $1 \leq i \leq T$ and let

$$\sum_{i=1}^T A(i) \leq A, \quad \sum_{i=1}^T B(i) \leq B$$

Then, there exists $i = i_0$ such that

$$A(i_0) \leq 2A, \quad B(i_0) \leq 2B$$

Proof Let N_a be the number of integers i such that $A(i) > 2A$ and N_b the number of integers i such that $B(i) > 2B$. Then,

$$\sum_{i=1}^T \frac{1}{T} A(i) > N_a \frac{2A}{T}$$

and $N_a < T/2$. Similarly $N_b < T/2$. The number of integers i such that $A(i) > 2A$ or $B(i) > 2B$ is bounded above by $N_a + N_b < T$. Therefore, the number of integers i such that $A(i) \leq 2A$ and $B(i) \leq 2B$ is bounded below by $T - N_a - N_b > 0$ and there exists an i_0 to which the theorem applies. Q.E.D.

Appendix B

Lemma Let the rate-distortion function $R(D)$ have a continuous first derivative and let $D(R)$ be the inverse of $R(D)$. Then, $F(R) = (1 - a/D(R))(R_{\max} - R) + R$ is decreasing in increasing R .

Proof The derivative of $F(R)$ is

$$F'(R) = \frac{a}{D^2(R)} [(R_{\max} - R) D'(R) + D(R)]$$

Since $D(R)$ is convex downward in R and $D(R_{\max}) = 0$, it follows that the bracketed term is bounded above by 0. Q.E.D.

References

1. C. E. Shannon, "Coding Theorems for a Discrete Source with a Fidelity Critereon," IRE National Convention Record, Part 4, pp. 142-163, 1959.
2. J. E. Savage, "The Complexity of Decoders-Part II: Computational Work and Deocding Time," IEEE Transactions on Information Theory, *A71-23102*
Vol. IT-17, No. 1, January, 1971.
3. J. T. Pinkston, "Encoding Independent Sample Information Sources," M.I.T. Research Lab. of Electronics, Cambridge, Mass., Tech. Rept. TR 462, October, 1967.
4. T. J. Goblick, "Coding for a Discrete Information Source with a Distortion Measure," Ph.D. thesis, Dept. of Electrical Engineering, M.I.T., Cambridge, October, 1962.
5. R. J. Pile, "The Transmission Distortion of a Source as a Function of the Encoding Block Length," Bell System Technical Journal, pp. 827-886, July-August 1968.
6. J. E. Savage, "Computational Work and Time on Finite Machines," paper in preparation.
7. O. W. Stalstrom, "The Complexity of Decoders for Iterated Hamming codes," Report, Division of Engineering, Brown University, August 1969.

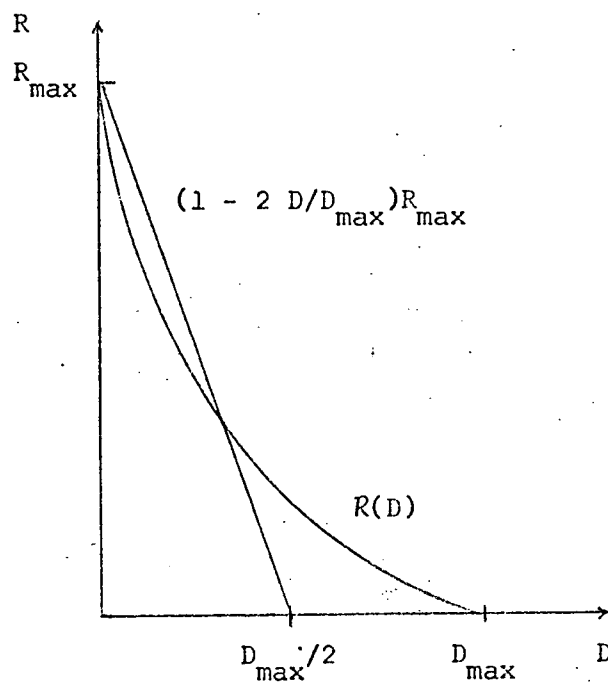


Fig. 1 Rate Distortion Function and Locus of Condition of Theorem 2